1989

Finding Point Correspondences and Determining Motion of a Rigid Object from Two Weak Perspective Views

Chia-Hoang Lee

Thomas Huang

Report Number:
88-748

FINDING POINT CORRESPONDENCES AND DETERMINING MOTION OF A RIGID OBJECT FROM TWO WEAK PERSPECTIVE VIEWS

Chia-Hoang Lee
Thomas Huang

CSD-TR-748
March 1988

Finding Point Correspondences and Determining Motion of a Rigid Object from two Weak Perspective Views

Chia-Hoang Lee

Department of Computer Sciences
Purdue University
West Lafayette, IN 47907

Thomas Huang

Coordinated Science Laboratory
University of Illinois
Urbana, IL 61801

ABSTRACT

Given two images of an n-point configuration which undergoes 3D rotation, translation and scaling: Our problems are (i) How can we match the corresponding points in the two images? Can all the possible mapping be found ? (ii) What underlying motions and associated depth components of these points could account for the two images ? (iii) Can the object be recovered uniquely? This formulation of the n-point problem is in the most general setting and does not assume attributes or features. A natural question to ask is whether an n-point problem is equivalent to a set of fewer-point problem. This paper presents a method which reduces an n-point problem to a set of 4-point problems. The effort of reduction takes $O(n)$ steps and it also takes $O(n)$ steps to construct all possible mappings of an n-point set from the solution to a 4-point problem. Other results include (1) Coplanarity condition of four points in two views. (2) Recovering the tilt direction of the rotational axis using four points in two views. (3) Recovering the scaling factor.

March 22, 1989

1. Introduction

Finding correspondences between primitive elements in two images (referred to as the correspondence problem) is a fundamental task in stereopsis and motion analysis. Featured-based (discrete) approaches in motion analysis [11][12][15] often assume that the correspondence of elements between consecutive frames has beeen established. The conventional methods lean toward separating the correspondence problem from the motion/structure problems [9]:

> Apparently, these two problems [the correspondence problem and the structure from motion problem] are solved independently by the human visual system... The critical empirical evidence for this is that none of the measurement on which the correspondence process rests involve three-dimensional angles or distance-- they are all two-dimensional measuremnets ... Thus, there is no deep need for any feedback from the latter task to the earlier.

With this viewpoint, the solutions for resolving correspondences among primitives lies in a fundamental principle which says the images of a given physical point should be similar to each other, at least in certain respects. For example, [4] uses sign change and orientation of the local zero-crossing contours as attributes of candidates, [3] uses phtometric information as attributes of an area, [8] attempts to characterize local structure, [10][14] suggest keeping track of features through small incremental movement.

Little thought was given on how the knowledge of an unknown motion itself might facilitate the correspondence process. To what extent, could properties of a rigid motion (unknown) be used to perform the task of matching without involving the attributes of the candidate points ? This turns the correspondence problem into a geometrical relationship among a collection of primitives in two frames. It should be emphasized that we do not advocate ignoring attributes or local structures. Rather, the viewpoint is that we first examine the correspondence process from motion constraints, and then the results derived under this general formulation could be used to supplement

the conventional methods.

In this paper we investigate with regard to weak perspective views how to relate the correspondence, the motion and the structure of an n-points object, given two images which might have different scales. Section 2 defines and discusses the problem. Section 3 develops our techniques and discusses advantage and disadvantage of using weak perspective projective projection model. Section 4 contains experiments and discussions.

2. Problem Statement

2.1 Problem Formulation

Consider an object consisting of n points denoted by A_i $(0 \le i \le n-1)$ in 3D space. Let $\overline{A_i}$ be the orthographic projection of A_i into the image plane. We can rotate, translate and scale the object and observe the effect on the projections of the n points in the image plane. A precise mathematical model is as follows: Let $B_i = \sigma\ (\mathbf{R}\ A_i + \mathbf{T}) = \sigma\ \mathbf{R}\ A_i + \sigma\ \mathbf{T}$ where $\mathbf{R}$ denotes 3D rotation; $\mathbf{T}$ denotes 3D translation; σ is a nonzero positive constant. The orthographic projection of B_i onto the image plane are denoted by $\overline{B_i}$. Thus we have two sets: FIRST $\equiv$ { $\overline{A_i} : 0 \le i \le n-1$; $\overline{A_i}$ is the first two components of A_i } and SECOND $\equiv$ { $\overline{B_i} : 0 \le i \le n-1$; $\overline{B_i}$ is the first two components of B_i }. Figure 1 depicts a situation in which FIRST consists of ten $\square$'s and SECOND consists of ten $\cdot$'s. The followings are our problems and will be referred to collectively as the n-point problem.

(i) How do we match the corresponding points in the two images? Can all possible mappings be found ?

(ii) What underlying motions and associated depth components of these points could account for the two images ?

(iii) Can the point configuration be recovered uniquely ?

The justification of introducing σ in the model comes from using orthographic or parallel projection to approximate perspective projection. It is well recognized that parallel projection is a reasonable approximation to perspective projection subject to a scale if the distance of the object to the camera is much larger than the size of the object. This same approximation was observed by Roberts [17], which he calls "weak" perspective. Mundy [16] also uses this approximation, which he calls affine transformation, and contains detail discussions:

> The affine approximation is usually quite reasonable for practical viewing application. Suppose one is imaging an industrial object 5 inches in diameter from 50 inches away. For an image sensor 0.5 inches square, the focal length required is about 2.5 inches, assuming that the object subtends half the field of view..... As another example, consider outdoor image applications such as obstacle recognition.... again the approximation is valid... The approximation is violated for cases ... A typical example .. is a scene showing a road .. extending off to the horizon.

The scale might be different in the two images due to a significant motion of an object (usually a large translational movement along the optical axis). This justification is reasonable for the case of the discrete approach but might not be for the case of continuous or flow-based approach. In the flow-based approach, the incremental motion between two consecutive frames is usually very small which implies that the scale remains more or less the same. However, the accumulation effects over a long period of time would result in a scale factor change.

2.2 Discussion

A more complicated problem is to consider a scene containing several objects, each undergoing its own motion. Thus FIRST and SECOND will contain projections of several mixed objects. Another generalization of the problem is to add uncertainty

into the model. For example, the positions of the projections may be corrupted by noise, or some points might appear in one image but not in the other due to occlusion or other reasons. Further, a curved segment or a region may replace a point in the problem.

For problem (i), there are n! different one to one mappings between the two images. The computational complexity of examining all mappings becomes intolerable as n becomes large. Further, there seems to be no systematic method of rejecting an inconsistent mapping. In general, a unique mapping between FIRST and SECOND does not exist. To see this, consider several points, uniformly separated around a circle in space, undergoing an arbitrary motion; then many mappings are possible. It is also apparent that a unique recovery for problem (iii) is generally not possible in this setting (no attributes or other structural properties). As a trivial example, consider the case of two points; then there are infinitely many objects which are consistent with the images. Also [9] note that

> With just two views, any number of points can be constructed that no unique three-dimensional interpretation (although some combination fortunately will)...

Our theory would systematically resolve the above Marr's argument (i.e "if some combination fortunately will" and uniqueness). Our approach to these problems comes from the concept of basis in linear algebra. The concept of basis in space leads us to ask if an *n*-point problem can be reduced to a set of fewer-point problems. There are two aspects of this question. The first aspect is the reduction step which enables one to study *n*-point problem on more manageable sets without affecting the answers. The second aspect is to analyze these problems on small sets. In section 3 we first discuss how to reduce a n-point problem to a 4-point problem. Next, we devote ourselves to analyzing the 4-point problem.

3. Theory

This section consists of three subsections. Subsection 3.1 describes the basic notation. Subsection 3.2 deals with constructing correspondences from four points to n points. We discuss how to establish all possible correspondences for the n-point problem. Subsection 3.3 discusses how to solve for all possible motions for a four-point problem.

One of the points, denoted by A_0 from FIRST will be chosen as the reference point and as the origin of the coordinate system. By doing this, the rotational axis is adjusted to pass through A_0 without changing the rotational matrix **R**. Further the translation **T** becomes zero if relative displacement instead of absolute position is used. The argument is summarized as follows: Let FIRST and SECOND be given, and let A_0 be the reference point. Using A_0 as the origin of the coordinate system, we have, for $1 \le i \le n-1$,

$$\begin{aligned} B_i - B_0 &= \sigma\,(\mathbf{R}\,A_i + \mathbf{T}) - \sigma\,(\mathbf{R}\,A_0 + \mathbf{T}) \\ &= \sigma\,\mathbf{R}\,(A_i - A_0) \end{aligned}$$

Renaming $B_i - B_0$ as B_i and $A_i - A_0$ as A_i, we have $B_i = \sigma\,\mathbf{R}\,A_i$ for $1 \le i \le n-1$ with the understanding that A_0 is also an object point. For the time being, the correspondence between A_i and B_i is given for the purpose of establishing several basic tools. Later this assumption will be removed.

3.1 Notation

$A_0 \equiv O$, A_1, A_2 -- three points in the scene

$\bar{A}_0 \equiv O$, $\bar{A}_1$, $\bar{A}_2$ -- projections of the above three points.

$B_0 \equiv O$, B_1, B_2 -- same three points in the scene after motion and scaling

$\bar{B}_0 \equiv O$, $\bar{B}_1$, $\bar{B}_2$ -- projections of the above three points.

Lower case Greek letters are used as scalars.

The rotational matrix denoted by **R** and its principal 2×2 minor $\mathbf{R}^*$ are written as

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = [r_1 \;\; r_2 \;\; r_3]\,; \;\; \mathbf{R}^* = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

where r_i is the ith column of **R**. We will denote $(r_{13}, r_{23})^t$ by cl_1 and $(r_{31}, r_{32})^t$ by cl_2 where t denotes transpose; l_1, l_2 are called the matching directions, and $||l_1|| = ||l_2|| = 1$; further $cl_1{}^*$ denotes $(-r_{23}, r_{13})^t$, $cl_2{}^*$ denotes $(r_{32}, -r_{31})^t$, i.e. $l_1{}^*$ is perpendicular to l_1, and $l_2{}^*$ is perpendicular to l_2.

It was shown in [12] that the rotational matrix has the following form:

$$\mathbf{R} = \begin{bmatrix} n_1{}^2 + (1-n_1{}^2)\cos\theta & n_1 n_2(1-\cos\theta) - n_3\sin\theta & n_1 n_3\,(1-\cos\theta) + n_2\sin\theta \\ n_1 n_2\,(1-\cos\theta) + n_3\sin\theta & n_2{}^2 + (1-n_2{}^2)\cos\theta & n_2 n_3\,(1-\cos\theta) - n_1\sin\theta \\ n_1 n_3\,(1-\cos\theta) - n_2\sin\theta & n_2 n_3(1-\cos\theta) + n_1\sin\theta & n_3{}^2 + (1-n_3{}^2)\cos\theta \end{bmatrix}$$

where $(n_1 \; n_2 \; n_3)$ is the rotational axis ; θ is the rotational angle; and the tilt direction of the rotational axis is given by $(n_1 \; n_2)$.

We shall call the following **R** as degenerate motion.

$$\mathbf{R} = \begin{bmatrix} \cdot & \cdot & 0 \\ \cdot & \cdot & 0 \\ 0 & 0 & \pm 1 \end{bmatrix}$$

This class of motions can arise in several ways:

(1) The optical axis is the rotational axis, i.e., every point rotates about the viewing axis.

(2) The direction of the rotational axis can be oriented arbitrarily, but the rotational

angle is a multiple of 360 degrees, which is equivalent to no motion.

(3) The rotational axis lies on the image plane and the rotational angle is 180 degrees.

The physical meaning of degenerate motion is clear if we treat the object as stationary with the relative motion in space attributed to the observer. In such cases, the projection plane remains the same for these frames. In other words, the observer does not change his viewing direction. Obviously, these frames can be generated through rotation or reflection or both from a single 2D image. The information are essentially equivalent to a single image with regard to the structure/motion of the object. Throughout the paper, the motion is assumed to be nondegenerate.

Whenever three points in two images are found (for instance $O, \bar{A}_1, \bar{A}_2$ are mapped to $O, \bar{B}_1, \bar{B}_2$), then it induces a mapping f from first image plane R^2 into second image plane R^2 as described below:

Definition:

$$f(\alpha\bar{A}_1 + \beta\bar{A}_2) = \alpha\, f(\bar{A}_1) + \beta f(\bar{A}_2)\,;\; f(\bar{A}_1) = \bar{B}_1\,;\; f(\bar{A}_2) = \bar{B}_2.$$

where α, β are real numbers.

O will be chosen appropriately so that $\bar{A}_1$, and $\bar{A}_2$ form a basis in R^2. f^{-1} will denote the inverse mapping by exchanging $\bar{A}_i$ and $\bar{B}_i$ (i = 1,2). The mapping f here is usually called an isomorphism between two algebraic structures.

3.2 Four Points vs N points

In this subsection, we first introduce the concept of matching direction (similar to epipolar line), which is useful for finding possible matches in the second frame when a point in the first frame is given. This is an important step, though straightforward. Next, we derive a coplanarity condition and show how to find four noncoplanar points. From four noncoplanar points, we use f (the induced mapping) to derive matching directions l_1, l_2 up to a scalar. From matching directions l_1, l_2 and f, we show the

possible correspondences of an image point A is restricted to the line of $f(\bar{A}) + d\,\boldsymbol{l}_1$ where d is a parameter. Furthermore, we show (lemma 3) that any point on $f(\bar{A}) + d\boldsymbol{l}_1$ in the second image, given the correspondence over the four points is realizable, can be taking as a corresponding point of $\bar{A}$ where $\bar{A}$ is any point in the first image. Lemma 3 provides a basis to extend the correspondence over a given set of four points to n points without affecting the possible underlying motions. A reduction algortihm to find all possible mappings with computation complexity $O(n^5)$ is then given.

Fact 1: Let **R** be a rotation and (a, b) be an image point in the first frame; then the coordinates of the corresponding image point in second frame are $(a\,r_{11} + b\,r_{12} + s\,r_{13},\ a\,r_{21} + b\,r_{22} + s\,r_{23})$ where s (a parameter) is the depth associated with (a, b). In other words, the corresponding point in the second frame can be any point of a line passing through $(a\,r_{11} + b\,r_{12},\ a\,r_{21} + b\,r_{22})$ in direction (r_{13}, r_{23}) i.e. $\boldsymbol{l}_1$.

Theorem 1: Let $O, \bar{A}_1, \bar{A}_2$ and $O, \bar{B}_1, \bar{B}_2$ be two projections of three points undergoing nondegenerate motion. Let $\bar{A}_3$ and $\bar{B}_3$ be a correspondence in the two images. Then a necessary and sufficient condition for O, A_1, A_2, A_3 to be coplanar in space is that $f\bar{A}_3 = \bar{B}_3$.

Proof: See Appendix 1.

Theorem 1 provides a simple coplanarity criterion for four points from two images. Actually, the information contained in the fourth point becomes redundant after coplanarity is recognized. One can see that the movement of any point lying on this plane can be easily established from available observables by f. Intuitively, the motion of three noncolinear points in a planar patch determines the movement of the patch in the image. However, the movement of a point which is not on this planar patch could not yet be determined.

Since theorem 1 is a necessary and sufficient condition, it implies that the condition $f(\bar{A}_3) \neq \bar{B}_3$ would enable one to conclude O, A_1, A_2, A_3 to be noncoplanar. In fact, lemma 2 below shows that other information could be deduced.

Lemma 1: Given two different weak perspective views of four noncoplanar points and the correspondence between them. Then the matching directions in the two views are uniquely determined (up to a sign) from the images. In fact, given the same notation as in theorem 1, $f(\bar{A}_3) - \bar{B}_3$ has the same direction as l_1 and $f^{-1}(\bar{B}_3) - \bar{A}_3$ has the same direction as l_2.

Proof: See Appendix 2.

Lemma 2: Let a correspondence between four noncoplanar points in the two frames and a fifth point $\bar{A}$ in the first image be given. Then a necessary condition for a point in the second image to be a correspondence of $\bar{A}$ is to lie on the line passing through $f(\bar{A})$ with direction l_1.

Proof: See Appendix 3.

Lemma 3: Let $\Omega = (\sigma, \mathbf{R}, \mathbf{T})$ be a motion which can account for a correspondence over four noncoplanar points in the two frames. Let a fifth point $\bar{A}$ in the first image be given. Let $\bar{B}$, which is in the second image, be any point on the line passing through $f(\bar{A})$ with direction l_1. Then Ω can always account for the extended correspondence by mapping $\bar{A}$ to $\bar{B}$.

Proof: See Appendix 4.

Recall that FIRST and SECOND both contain n image points. We try to answer systematically how many mappings are permissible, what are these mappings, what are possible underlying motions for each permissible mapping? Is there any situation which allows a unique motion recovery?

Suppose G is the set of motions which could account for a given correspondence m (i.e m is realizable) between FIRST and SECOND. Our goal is to find m and G. The approach is first to identify four noncoplanar points $\{O, \bar{A}_1, \bar{A}_2, \bar{A}_3\}$ in the first image. Next we have to consider all possible mappings from $\{O, \bar{A}_1, \bar{A}_2, \bar{A}_3\}$ into SECOND because one of the goals is to find out all permissible mappings. There are $O(n^4)$ mappings to be examined. Obviously m restricted to $\{O,\bar{A}_1,\bar{A}_2,\bar{A}_3\}$ is one of the $O(n^4)$ mappings. Once this restricted m is under examination, one would find out that $m(\bar{A}_i)$ lies on the line of $f(\bar{A}_i) + d_i\, l_1$ for $4 \le i \le n-1$. Using lemma 3, one could derive m from the restricted m. For G, which is the set of motions associated with m, one knows that G is related to the four points from lemma 3. We will discuss how to solve for it in the next section. For the time being, we will assume that all $O(n^4)$ mappings are realizable.

However, there is no way to identify four noncoplanar points $\{O, \bar{A}_1, \bar{A}_2, \bar{A}_3\}$ from a single image unless other information and other techniques (e.g. shape from shading, shape from shape, or structural properties) are used. One possible method using two images is simply to choose a subset of four points from FIRST and considers n(n-1)(n-2)(n-3) possible mappings into SECOND. Theorem 1 then tells us which mapping interprets the four points to be noncoplanar. For these cases, one could proceed to extend correspondence as discussed above. For those coplanar cases, one has to consider extra point in order to break the coplanarity. This approach would require at least complexity of $O(n^4)$ to begin with.

An alternative to searching for four noncoplanar points is to start with three instead of four points. First we choose any three noncollinear points (name them as $O,\bar{A}_1,\bar{A}_2$ from FIRST and consider all possible mappings into SECOND. This gives rise to n(n-1)(n-2) possible (not necessarily permissible) ones. We will use $O,\bar{B}_1,\bar{B}_2$ as the name of corresponding elements in SECOND for each mapping. As mentioned before, an f is induced between the two image planes for each mapping. We next per-

form the test $f(\bar{A}_i) \in SECOND$ for $3 \le i \le n-1$. If one of the test (say $i = k$) is not true, then $\bar{A}_k$ is the fourth point to be added to the original three points. Since no matter how $\bar{A}_k$ is mapped into SECOND (there are n-3 possibilities), one always interprets $O,\bar{A}_1,\bar{A}_2,\bar{A}_k$ to be noncoplanar. If tests are true for every i, then there are (n-3)(n-4)/2 + 1 mappings to be generated from three points to find out the existence of four noncoplanar points. Among them (n-3)(n-4)/2 mappings have interpretation of four noncoplanar points, and one of them has found a planar patch interpretation for these n points. To see this, one notices that, for the f under consideration, $f(\bar{A}_i) \in SECOND$ are true for $3 \le i \le n-1$. Thus one obtains a coplanar interpretation by sending $\bar{A}_i$ to $f(\bar{A}_i)$ for $3 \le i \le n-1$. Now which point does one have to include into the set of $\{O,\bar{A}_1,\bar{A}_2\}$ to have four noncoplanar points? Clearly, one could include $\bar{A}_3$ by not mapping it to $f(\bar{A}_3)$. Consequently, there are n-4 mappings here. However, the coplanar mapping which sends $\bar{A}_3$ to $f(\bar{A}_3)$ could not be ignored (this means $\bar{A}_3$ is bound now). Thus one has to consider including $\bar{A}_4$ into the set, which would yield (n-5) mappings for noncoplanar interpretation. Continuing in this manners, one obtains (n-3)(n-4)/2 mappings for existence of four noncoplanar points and one mapping for planar patch interpretation.

The algorithm to extend G3F = $\{O,\bar{A}_1,\bar{A}_2\}$ to four-element subset is given below:

Algorithm:

(*Initialization*):

Let $\{\bar{A}_3 \cdot\cdot \bar{A}_n\}$ = FIRST - G3F;
Create n(n-1)(n-2) $\alpha_i's$ (mappings from
G3F into SECOND) and let
$ALPHA = \{\alpha_i : 1 \le i \le n(n-1)(n-2)\}$

for $\alpha_i \in ALPHA$ **do**
Q_{α_i} = SECOND - α_i(G3F); /*Q_{α_i} consists of $\bar{B}_3 \cdot\cdot \bar{B}_n$ /*

for $3 \le j \le n$ **repeat**
$R_j := f_{\alpha_i}(A_j)$
until $R_j \in Q_{\alpha_i}$ is false;

case:

$j \ne n-3$: /* spawn n-3 mappings */
Add A_j to G3F and spawn n-3 mappings
by sending A_j to each element in
SECOND - α(G3F) respectively;

$j = n-3$: /* spawn $(n-3)(n-4)/2 + 1$ mappings */
Record the "coplanar mapping" which sends A_i to R_i ;

Add $\bar{A}_3$ to G3F and spawn n-4 mappings by
sending A_3 to each element in
SECOND - α_i(G3F) – $f_{\alpha_i}(\bar{A}_3)$ respectively;

Spawn n-5 mappings by sending $\bar{A}_3$
to $f_{\alpha_i}(A_3)$ and A_4 to n-5 elements of
SECOND - α_i(G3F) – $f_{\alpha_i}(\bar{A}_3)$ – $f_{\alpha_i}(\bar{A}_4)$ respectively;

Continuing in this manner, one spawns (n-4) + (n-5) + ...+1 mappings.

end

The computations needed can be summarized as follows: First, choose a subset G3F of three noncollinear points from FIRST. Construct $n(n-1)(n-2)$ mappings from G3F into SECOND. Second, for each one of $n(n-1)(n-2)$ mappings, one needs to extend G3F to G4F. It might generate $n-3$ or n^2 mappings. The computation required to generate n-3 or n^2 mappings might take 1 step or n-3 steps. Third, for each mapping from G4F into SECOND, one applies lemma 3 to extend or reject the mapping. This

requires $n-3$ step. Consequently, one needs computations $O(n^5)$ roughly to search for all permissible mappings between FIRST and SECOND, assuming each mapping over four points is realizable.

3.3 Solutions for four points problem

This section shows any arbitrary mapping between given four points in two images is always realizable. Next we show how to decide all possible motions for a given mapping and thus the relative positions of a n-points object.

Theorem 2: Given a mapping between two sets of four noncoplanar points in two frames, the scale σ can be recovered uniquely. In fact $\bar{B}_1 \cdot l_1^* = \sigma \bar{A}_1 \cdot l_2^*$

Proof: See Appendix 5.

The implication of theorem 2 is twofold: First it shows a benifit gained by using weak perspective to approximate perspective projection. Imagine an object (with four noncoplanar featured points) undergoing an unknown motion. Three models could be used for problem formulation. The first model is to use perspective projection; the second one is to use parallel projection; the third one is to use weak perspective model. In the third model, theorem 2 gives us approximately how far the object travels away from the camera along the optical direction. Second it leads us to decide in the next theorem what are the possible correspondences between the two set of four points in two images. Further it enable (see theorem 4) us to determine a one parameter family of possible motions explictly.

Examine the following formula:

$$\bar{B}_1 \cdot l_1^* = \sigma \bar{A}_1 \cdot l_2^{*\#} \tag{3.3.1}$$

$$\bar{B}_2 \cdot l_1^* = \sigma \bar{A}_2 \cdot l_2^{*\#} \tag{3.3.2}$$

$$\bar{B}_3 \cdot \boldsymbol{l}_1^{\,*} = \sigma\, \bar{A}_3 \cdot \boldsymbol{l}_2^{\,*\#} \tag{3.3.3}$$

If one adds these three equations multiplied by proper coefficients, respectively, one obtains

$$(f(\bar{A}_3) - \bar{B}_3) \cdot \boldsymbol{l}_1^{\,*} = 0$$

$$(f^{-1}(\bar{B}_3) - \bar{A}_3) \cdot \boldsymbol{l}_2^{\,*} = 0$$

Equations (3.3.1-3.3.3) describe how $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ (up to a unknown constant) are derived. Next one substitue $\boldsymbol{l}_1^{\,*}$ and $\boldsymbol{l}_2^{\,*}$ back to any one of the equations to derive σ. Thus these three equations are always consistent when σ is treated as a parameter (one degree of freedom). However, if σ is taken to be 1 as it is in the pure parallel projection model, then these three equations are not necessarily consistent. The following theorem shows that any noncoplanar mapping of four points is always realizable (i.e. accountable by some rigid motion) in weak perspective model. In fact, theorem 4 gives a one parameter family of motions explictly for each noncoplanar mapping.

Theorem 3: Any mapping of four points in the two images can always be realized.

Proof: Recall

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}^* & c\boldsymbol{l}_1 \\ c\boldsymbol{l}_2^{\,t} & r_{33} \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & r_{13} \\ \cdot & \cdot & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}; \; c\,\boldsymbol{l}_1^{\,*} = \begin{pmatrix} -r_{23} \\ r_{13} \end{pmatrix}; \; c\,\boldsymbol{l}_2^{\,*} = \begin{pmatrix} r_{32} \\ -r_{31} \end{pmatrix} \quad \|\boldsymbol{l}_1\| = \|\boldsymbol{l}_2\| = 1.$$

Whether a given mapping could be realized or not, it depends on the solutions to the equations below.

$$\bar{B}_1 = \sigma\, \mathbf{R}^*\, \bar{A}_1 + \sigma\, s_1\, c\, \boldsymbol{l}_1$$

$$\bar{B}_2 = \sigma\, \mathbf{R}^*\, \bar{A}_2 + \sigma\, s_2\, c\, \boldsymbol{l}_1$$

$$\bar{B}_3 = \sigma\, \mathbf{R}^*\, \bar{A}_3 + \sigma\, s_3\, c\, \boldsymbol{l}_1$$

It is clear that σ, l_1 and l_2 could be decided. However c is a unknown constant between -1 and 1. The task here is to find solutions for s_1, s_2, s_3 and $\mathbf{R}^*$ such that $\mathbf{R}^*$ is the principle minor of a rotational matrix. Since the third column of $\mathbf{R}$ is perpendicular to other columns, we get

$$l_1{}^t \mathbf{R}^* = -r_{33}\, l_2{}^t$$

and from [5]

$$l_1{}^{*t} \mathbf{R}^* = l_2{}^{*t}$$

Rewriting them into matrix form, one gets

$$\begin{bmatrix} l_1{}^t \\ l_1{}^{*t} \end{bmatrix} \mathbf{R}^* = \begin{bmatrix} -r_{33} l_2{}^t \\ l_2{}^{*t} \end{bmatrix}$$

Thus

$$\mathbf{R}^* = \begin{bmatrix} l_1{}^t \\ l_1{}^{*t} \end{bmatrix}^{-1} \begin{bmatrix} -r_{33} l_2{}^t \\ l_2{}^{*t} \end{bmatrix}$$

$$= (l_1 \;\; l_1{}^*) \begin{bmatrix} -r_{33} l_2{}^t \\ l_2{}^{*t} \end{bmatrix} = (l_1 \;\; l_1{}^*) \begin{bmatrix} -\sqrt{1-c^2}\, l_2{}^t \\ l_2{}^{*t} \end{bmatrix}$$

The constructed $\mathbf{R}^*$ above indeed makes $\mathbf{R}$ a rotation (one could check it by writing it out). Notices that c is the only parameter. To show that s_1, s_2, s_3 can be found subsequently, one has to demonstrate that $\bar{B}_i - \sigma \mathbf{R}^* \bar{A}_i$ is indeed parallel to l_1. Writing it out, one gets

$$\bar{B}_i - \sigma \mathbf{R}^* \bar{A}_i = \bar{B}_i - \sigma (l_1 \;\; l_1{}^*) \begin{bmatrix} -\sqrt{1-c^2}\, l_2{}^t \\ l_2{}^{*t} \end{bmatrix} \bar{A}_i$$

$$= \bar{B}_i - \sigma (l_1 \;\; l_1{}^*) \begin{bmatrix} -\sqrt{1-c^2}\, l_2{}^t \bar{A}_i \\ l_2{}^{*t} \bar{A}_i \end{bmatrix}$$

$$= \bar{B}_i - \sigma [\, -\sqrt{1-c^2}\, l_2{}^t \bar{A}_i\, l_1 + l_2{}^{*t} \bar{A}_i\, l_1{}^*]$$

It now suffices to show $\bar{B}_i - \sigma\, \boldsymbol{l}_2^{*t}\, \bar{A}_i\, \boldsymbol{l}_1^{*}$ is parallel to $\boldsymbol{l}_1$. For this, we show it is perpendicular to $\boldsymbol{l}_1^{*}$:

$$(\bar{B}_i - \sigma\, \boldsymbol{l}_2^{*t}\, \bar{A}_i\, \boldsymbol{l}_1^{*}) \cdot \boldsymbol{l}_1^{*}$$

$$= \bar{B}_i \cdot \boldsymbol{l}_1^{*} - \sigma\, \boldsymbol{l}_2^{*t} \bar{A}_i \quad (\boldsymbol{l}_1^{*} \cdot \boldsymbol{l}_1^{*} = 1)$$

$$= \bar{B}_i \cdot \boldsymbol{l}_1^{*} - \sigma\, \boldsymbol{l}_2^{*t}\, \bar{A}_i$$

$$= \bar{B}_i \cdot \boldsymbol{l}_1^{*} - \sigma\, \bar{A}_i \cdot \boldsymbol{l}_2^{*}$$

$= 0$ (see 3.3.1 –3.3.3) **Q.E.D.**

Theorem 3 gives a clear distinction between the weak perspective model and the parallel projection model concerning if a mapping is permissible or not. Using our model, any noncoplanar mapping is realizable and can be accounted for by a parameter family of motions. If a pure parallel projection model is used, then the given mapping could be realized only if the derived σ equals to 1.

The construction of $\mathbf{R}^*$ in fact is a representation of a general $\mathbf{R}$. One could see that $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ represent two degrees of freedom (remember their norms are both unity), and scalar c which is between -1 and 1 represents another degree of freedom.

4. Discussion

Our algorithm is based on the searching for four points in one image and examines all possible mappings of these four points into another image. There are roughly $O(n^4)$ mappings. For each mapping, one could derive the matching directions $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ from lemma 1. Once $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ are derived, theorem 3 gives a one parameter family of rotation and theorem 2 gives the scale. Recall that each mapping also induces a mapping f. Using f, one could derive a matching line $f(\bar{A}_i) + d_i\, \boldsymbol{l}_1$ for each $\bar{A}_i$. Next one would use lemma 3 to extend the mapping until it include all of n points.

Our algorihtm takes $O(n^5)$ to solve the n-points problem. This means (i) find out all realizable mappings (ii) find out all possible motions for each permissible mapping (iii) find out the relative depths with regard to each motion. An assumption is that all points are visible in each view. It is evident that several issues such as missing points, multiple objects, sensitivity, range of applicability of our model, and realistic applications might arise. We will discuss these issues briefly and give some simulation results.

4.1 Missing points and Multiple objects

If some of the points may be missing, then the chosen four points in one image may not be present in another image, while in the case of multiple objects, the chosen four points may not lie in the same object. Thus it becomes necessary to examine n(n-1)(n-2)(n-3) different subsets of four points from each image and thereby makes the computations $O(n^9)$ (recall that it takes $O(n)$ computations for each mapping). However, difficulties and complications exist in deciding the permissibility of a mapping. In the case of missing points a legitimate mapping does not necessarily find a correspondence for every point in the first image. It depends on how many points disappear. In the case of multiple objects, the knowldge of the number of objects is useful for deciding the possible members of objects (see case 3 in experimental section).

4.2 A Realistic Approach

In most practical problems, many attributes and local structures of primitive elements are present. Furthermore multiple objects and missing points are likely to occur. In general there is no need to search for all possible mappings which could be theoretically realized. What it requires is to search for the one which already exists. It goes without saying that a good vision system should use as much information as possible for the purpose of robustness. Starting with $O(n^4)$ or $O(n^8)$ mappings to perform matching is an enormous task in our algorithm. On the other hand conventional methods are based on a good heuristic principle - the images of a given physical point

should be similar to each other. Thus a good approach is to combine our results with conventional matching methods. First one could use visible attributes or local structures to perform a partial matching. Second, our technique (requiring four points) could be used to predict the locations of matching for other points. Third one could use conventional method to perform matching along these predicted locations. By doing this, one provides a mathematical basis to supplement the conventional techniques.

5. Simulation Results

Three simulation experiments to test the analysis were performed on a VAX/8600. In the first simulation image data were generated from weak perspective model. The purpose is to demonstrate the theoretical aspects of the technique. In the second simulation image data were generated from a perspective model. Furthermore, various amounts of noise ranging from one pixel to three pixels were added to the data. The goal is to study how often our algorithm succeeds in finding the correct mapping; and to study the effect of using a weak perspective model to approximate a perspective model. In the third simulation image data were generated (using weak perspective) from two objects each exercising different motions. The aim is to illustrate the issues related to multiple objects.

Simulation 1(a): Reduction Algorithm

The experiment to demonstrate the reduction algorithm is as follows: Consider the two images in Figure 1: FIRST = set of ten □'*s* and SECOND = set of ten ·'*s*. We first choose 3 □'*s* and denote them by G3F. Clearly there are 10×9×8 = 720 possible mappings into SECOND. We apply the reduction algorithm to search for a forth point so that it would extend G3F into four noncoplanar points under each of these mappings. A simple program can be written using three nested *for loops* with indexes from 1 to 10, but requiring that no two indexes be the same. We choose the first three □'*s* as G3F

and the first one is the reference point i.e. O of the first image. For each of the 720 mappings, we start to examine if $f(\bar{A}_i)$ is a member of SECOND for $4 \leq i \leq n$ where f is an induced mapping. Surprisingly, $f(\bar{A}_4)$ does not belong to SECOND for any possible mappings (recall that SECOND changes if the reference point is chosen differently in the second image). Thus G3F is expanded to include $\bar{A}_4$ and spawn 7 possible mappings due to inclusion of $\bar{A}_4$. Now we have a total of $720 \times 7 = 5040$ mappings. For each one of them, the extension to 10 $\square$'s is attempted. Only one mapping (the correct one and unique in this example) could be extended to include all of the $\square$'s. There are three mappings which could be extended to one more point. The rest of 5036 mappings could not be extended at all.

Simulation 1(b): Use apriori information in scale to reduce searching

Considering the above 5040 mappings from $\bar{A}_1, \bar{A}_2, \bar{A}_3, \bar{A}_4$ into SECOND. If the relative scale between the two images is believed *a priori* to be between 0.98 and 1.02, then 70 mappings of which are realizable. However, none of these could be extended over all n points. If the relative scale between the two images is *a priori* between 0.99 and 1.01, then 32 mappings of which can be realized. Again none of these 32 mappings could be extended over n points. Only one mapping of which the relative scale is 0.8 could be extended to include n points. In other words, there is a unique mapping for the input of two images depicted in Figure 1.

Simualtion 2: Data from perspective projection

Various experiments were run to investigate the sensitivity of our algorithm. Of all permissible solutions in any example, one is always the set of motion and depth parameters we used to generate the observables. The goal of simulations here is to study how often our algorithm could succeed in finding the correct mapping which arises from the set of parameters we have used. As for motion parameters, only tilt could be uniquely recovered based on weak perspective model while there are infinite pairs of slant and

rotational angle. This could be understood since only the matching directions l_1 , l_2 and scale σ could be recovered using this model. Furthermore $l_1 + l_2$ gives the tilt direction. Thus the results here would include (i) whether the correct mapping we have used could be found (ii) the tilt parameter of the rotational motion (iii) the scale between the two images.

The input data were generated using perspective projection in accordance with the following scenario: One is imaging an object 5 inches in diameter from 50 inches away. The image sensor is 0.5 inches square; the focal length is about 2.5 inches; and the object subtends less than half the field of view.

In the first experiment no noise were added to the input data. The goal is to investigate if a weak perspective model is a good approximation to a perspective model. We first choose any four points and manually establish their correspondence. Our algorithm correctly finds out the correspondence of the remaining points.

In the second series of experiments noise ranging from one pixel to three pixels were added to the input data. Again we choose four points and manually establish their correspondence. From these four points, our technique would first derive the motion parameters and matching directions. For each point in the first image a matching line in the second image would be derived. We then check the distance of the corresponding point to this matching line. We also derive the tilt and the scale. Totally 432 experiments are conducted. We found that the average distance of corresponding point to its matching line is 2 pixels and the standard deviation is 2 pixels also. The average scale is 1.20 which falls in the range between 1.16 and 1.22 calculated from each point of the object in the two scenes. The error in tilt angle is 45 degrees. From these, it seems that weak perspective is a good approximation for the recovery of scale and correspondence but not a good approximation model for recovery of the motion parameters.

Simulation 3: Multiple objects

Two objects, one containing 7 points, and the other 6 points, undergoing different motions, were generated using our model. Now, one could not simply choose a particular set of four points from the first image because object points are fused. All possible four points in the two images need to be examined. This gives roughly $O(n^8)$ to begin with. Since $O(n^8)$ is an enormous number for n = 13, we identify four points in each object from the first image. Then we consider all possible mappings from the chosen four points into SECOND. For the first set of four points, we have 17160 mappings to examine. Out of these, one mapping could be extended to cover 7 points. For the second set of four points, we again have 17160 mappings to examine. Out of these, one mapping could be extended to cover 7 points. The mapping, covering 7 points, generated from the first object shares one point with the mapping, covering 7 points, generated from the second object. Thus one has an ambiguity about classifing this point.

References

1. Aloimonos J. (1986), "Perception of Structure from Motion", Conference of Computer Vision and Pattern Recognition, IEEE Computer Society, June 25-27, Miami, Florida.
2. Aloimonos J. and A. Papageorgiou (1987), "On the Kinetic Depth Effect: Lower bounds, regularization and learning", CAR-TR-261, Center for Automation Research, University of Maryland.
3. Baker (1981), "Depth from Edge and Intensity Based Stereo", Ph.D thesis, AIM-347, Computer Science Department, Stanford University.
4. Grimson W.E.L (1981), *From Images to Surfaces*, (Cambridge: MIT Press).
5. Hunag, T.S (1986), "Motion and Structure from Two Orthographic Projections" Technical Report, Coordinated Science Lab. University of Illinois.
6. Jenkin M. and P.A. Kolers (1986), "Some Problems with Correspondence", RBCS-TR-86-10, Department of Computer Science, University of Toronto.
7. Nagel H.H. and B. Neuman (1981), "On 3-D reconstruction from two perspective views", *Proc. IJCAI 81*, Vol II.
8. Leclerc, Y.G. and S.W Zucker (1987), "The local Structure of Image Discontinuities in One Dimension", *IEEE Trans. PAMI 9*, 341-355.
9. D. Marr (1982), *Vision*, San Francisco, CA: Freeman.
10. Nevatia R. (1976), "Depth from Camera Motion in a Real World Scene", *Computer Vision, Graphics & Image Processing 9*, 203-214.
11. Roach, J.W. and J.K. Aggarwal, "Determining the movement of objects from a sequence of images", *IEEE Trans. PAMI 2*,
12. Rogers D.F. and J.A. Adams (1976), Mathematical Elements for Computer Graph-

ics, McGraw-Hill, New York.

13. Tsai, R. Y. and T.S. Huang (1984), "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces", *IEEE Trans. PAMI 6,* 13-27.

14 Williams , T.D. (1980), "Depth from Camera Motion in a Real World Scene", *IEEE Trans. PAMI 2,* 511-516.

15. Ullman, S. (1979), "The Interpretation of Visual Motion", MIT press.

16. Thompson D.W and J.L. Mundy (1987), "Three-Dimensional Model Matching From an Unconstrained Viewpoint", IEEE International Conference on Robotics and Automation, Raleigh, North Carolina

17. Robert L.G (1965), "Machine perception of three dimensional solids", in *Optical and Electro-optical Information profession,* J.T. Tippett et al. (Eds.), Cambridge, MA MIT Press.

Appendix 1

Proof of Theorem 1: The span, $\bar{A}_3 = \delta \bar{A}_1 + \gamma \bar{A}_2$, can always be assumed. To establish sufficiency, recall that $\mathbf{R}^*$ is a principal minor of $\mathbf{R}$. Since $B_i = \sigma R A_i$, assuming that s_i is the depth of A_i , we have

$$\bar{B}_i = \sigma R^* \bar{A}_i + \sigma s_i l_1 \qquad \text{for } 1 \le i \le 3 \tag{1}$$

Write $\bar{A}_3 = \delta \bar{A}_1 + \gamma \bar{A}_2$. Applying σR^* to this relation and using (1), we obtain

$$\sigma R^* \bar{A}_3 = \sigma \delta R^* \bar{A}_1 + \sigma \gamma R^* \bar{A}_2 \tag{2}$$

$$= \delta (\bar{B}_1 - \sigma s_1 l_1) + \gamma (\bar{B}_2 - \sigma s_2 l_1)$$

$$= \delta \bar{B}_1 + \gamma \bar{B}_2 - (\sigma \delta s_1 + \sigma \gamma s_2) l_1$$

Substituting (1) for the left hand side of (2), we have

$$\bar{B}_3 - \sigma s_3 l_1 = \delta \bar{B}_1 + \gamma \bar{B}_2 - (\sigma \delta s_1 + \sigma \gamma s_2) l_1$$

Hence

$$\sigma (\delta s_1 + \gamma s_2 - s_3) l_1 = \delta \bar{B}_1 + \gamma \bar{B}_2 - \bar{B}_3 = f(\bar{A}_3) - \bar{B}_3 \tag{3}$$

and

$$(\delta s_1 + \gamma s_2 - s_3) l_1 = 0\# \tag{4}$$

Since $\mathbf{l}_1$ is not a zero vector due to the assumption of nondegenerate motion, we have $(\delta s_1 + \gamma s_2 - s_3) = 0$. Hence, $A_3 = \delta A_1 + \gamma A_2$ whence coplanarity follows.

To show necessity, it is sufficient to observe that, since O, A_1, A_2, A_3 are coplanar, $\bar{A}_3 = \delta \bar{A}_1 + \gamma \bar{A}_2$ can be extended to $A_3 = \delta A_1 + \gamma A_2$. Now, applying rotation $\mathbf{R}$ and scale σ to the relation, we have $\delta B_1 + \gamma B_2 = B_3$. Consequently, we see that $f(\bar{A}_3) = \delta \bar{B}_1 + \gamma \bar{B}_2 = \bar{B}_3$. **Q.E.D.**

Appendix 2

Proof of Lemma 1: As seen in equation (3) of proof of Theorem 1, $(\delta s_1 + \gamma s_2 - s_3)\mathbf{l}_1 = f(\bar{A}_3) - \bar{B}_3$ holds. Since these four points are not coplanar, we have $\mathbf{f}(\bar{A}_3) - \bar{B}_3 \neq 0$. Thus we can derive $\mathbf{l}_1$ up to a sign (recall that l_1 is a unit vector. Furthermore, we can apply the same technique to derive $\mathbf{l}_2$ up to a sign by *interchanging the roles of the two frames.* **Q.E.D.**

Appendix 3

Proof of Lemma 2: Let A be any non-feature point with s as its depth; let B be the point corresponding to A after the motion. Write $\bar{A} = \delta\bar{A}_1 + \gamma\bar{A}_2$. Applying $\mathbf{R}^*$ and scale σ to the above relations, we have

$$\sigma \mathbf{R}^* \bar{A} = \sigma\delta \mathbf{R}^* \bar{A}_1 + \sigma\gamma \mathbf{R}^* \bar{A}_2 \quad \text{and} \quad \bar{B} = \sigma \mathbf{R}^*\bar{A} + \sigma s \boldsymbol{l}_1$$

Hence

$$\bar{B} = \delta(\bar{B}_1 - \sigma s_1 \boldsymbol{l}_1) + \gamma(\bar{B}_2 - \sigma s_2 \boldsymbol{l}_1) + \sigma s \boldsymbol{l}_1$$

Rearranging terms, we have

$$\bar{B} = \delta\bar{B}_1 + \gamma\bar{B}_2 + \sigma(s - \delta s_1 - \gamma s_2)\boldsymbol{l}_1 = f(\bar{A}) + \sigma(s - \delta s_1 - \gamma s_2)\boldsymbol{l} \quad (6)$$

From this equation, we know that $\bar{B}$ must lie on a line passing through $\mathbf{f}(\bar{A})$ with direction $\mathbf{l}_1$. **Q.E.D.**

Appendix 4

Proof of Lemma 3: Let $\bar{A} \in$ FIRST and $\bar{C}$ be possible correspondence of $\bar{A}$. From above lemma and equation (6), we know that

$$\bar{C} = f(\bar{A}) + \sigma(s - \delta s_1 - \gamma s_2)\boldsymbol{l}_1$$

From assumption of $\overline{B}$, we know that

$$\overline{B} = f(\overline{A}) + b\, \boldsymbol{l}_1$$

Since σ and $\mathbf{l}_1$ are known, we have $s - \delta\, s_0 - \gamma\, s_1 = b/\sigma$. Thus s can be derived. Therefore, if $\delta\, s_0 + \gamma\, s_1 + b/\sigma$ is assigned as the depth of $\overline{A}$, the same motion (σ,**R**,**T**) can account for the correspondence of $\overline{A}$ and $\overline{B}$. **Q.E.D.**

Appendix 5

Proof of Theorem 2: Recall that $\overline{B}_1 = \sigma\, \mathbf{R}^*\, \overline{A}_1 + \sigma\, s_1\, \boldsymbol{l}_1$ holds. Taking the scalar product of both sides with $\boldsymbol{l}_1{}^*$, one gets

$$\begin{aligned}
\overline{B}_1 \cdot \boldsymbol{l}_1{}^* &= \sigma \mathbf{R}^*\, \overline{A}_1 \cdot \boldsymbol{l}_1{}^* \\
&= \sigma \boldsymbol{l}_1{}^{*t} \mathbf{R}^*\, \overline{A}_1 \\
&= \sigma \boldsymbol{l}_2{}^{*t}\, \overline{A}_1 \\
&= \sigma\, \overline{A}_1 \cdot \boldsymbol{l}_2{}^*
\end{aligned}$$

Since $\mathbf{l}_1, \boldsymbol{l}_2$ can be derived up to a sign, thus σ can be derived. **Q.E.D.**

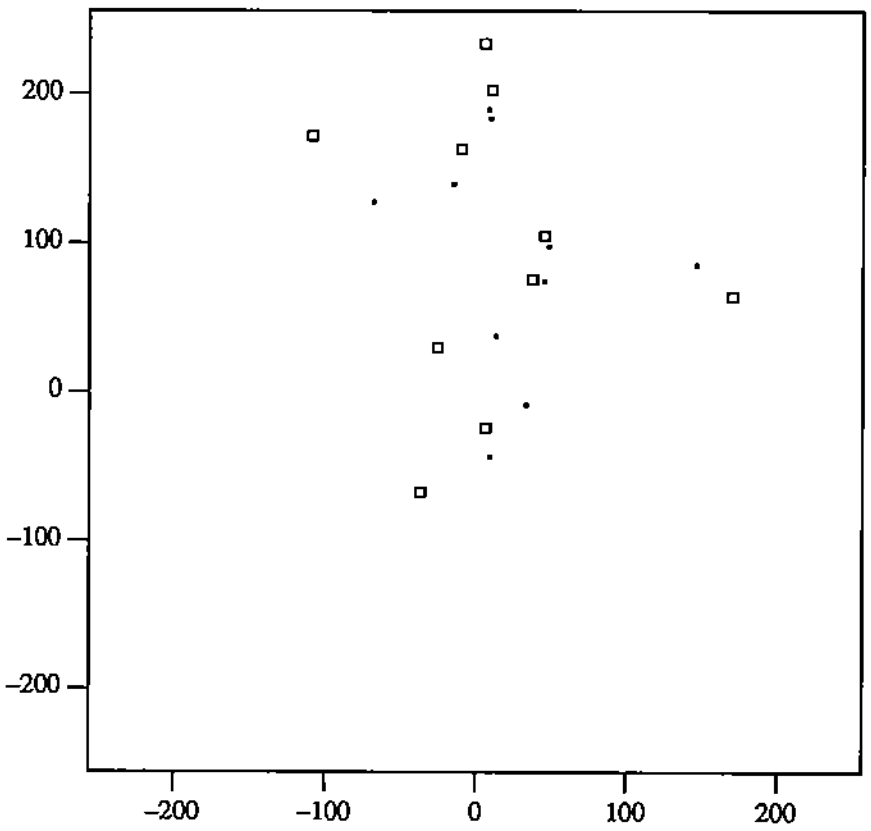

Figure 1. Squares represent the first image while dots represent the second image